DIGITAL SIGNAL PROCESSING **6,** 29–36 (1996)
ARTICLE NO. 0004

Porayath, R., J. F. Doherty, and S. F. Russell, "An Adaptive Echo Canceler for Hands-Free Teleconferencing," *Digital Signal Processing Journal*, Academic Press, vol. 6, no. 1, pp. 29-36, Jan.1996.

An Adaptive Acoustic Echo Canceler for Hands-Free Teleconferencing

Rajiv Porayath, John F. Doherty, and Steve F. Russell

Department of Electrical Engineering and Computer Engineering,
Iowa State University, Ames, Iowa 50011

Porayath, R., Doherty, J. F., and Russell, S. F., An Adaptive Acoustic Echo Canceler for Hands-Free Teleconferencing, *Digital Signal Processing* **6** (1996), 29–36.

A new technique for high quality acoustic echo cancellation in teleconference and speaker phone systems has been developed. The new echo canceler is based on an enhanced adaptive filtering algorithm and uses a spread-spectrum-based technique to mitigate the effects of double talk. The test facility used to develop this new acoustic echo canceler and the system design of echo canceler are described. © 1996 Academic Press, Inc.

I. INTRODUCTION

New telecommunications technology known as the "information superhighway," which will provide the capability for reasonable-cost two-way audio and video conferencing, is being rapidly implemented. With this capability, two geographically separate groups of people can meet in a "virtual" way and avoid the expense of travel. Although technology is making remote audio/video conferencing practical, the meeting rooms used for these conferences are still frequently plagued by acoustic echoes that make it difficult to carry on a meaningful conversation. The efficient mitigation of these deleterious effects in a hands-free telephony system is an active research and development area. The hands-free environment is characterized by human communicators that are not ordinarily colocated with their respective telephones. The decoupling of the handset from the head introduces the room as another element in the voice system. The result is multiple reflections of the transmitted acoustic energy impinging on the listening apparatus, which are transmitted back as delayed replicas of the speaker's voice. The nature of the acoustic echo patterns can change rapidly due to variations in the environment, for example, opening and closing of doors. The standard approach to eliminating the acoustical echo is to use a discrete-time linear adaptive filter at the receiver where the echo is predominant. This filter is used to quantitatively characterize the acoustical link between the speaker and the microphone, thus generating an electronically synthesized replica of the acoustical echo. This replica is then subtracted from the echo received by the microphone to decouple the speaker and the microphone. Figure 1 shows the standard approach used in acoustic echo cancellation. A major drawback in this technique is that the addition of near-end speech into the echo return path is not accounted for by the adaptive filter. Complicated and often marginally effective methods are used to alleviate this problem. One example is near-end speech detection followed by echo cancellation cessation. Another problem is the convergence rate of the adaptive filter. For a typical teleconference setup, the filter is on the order of 2000 taps and the input is correlated speech signals. This results in slow convergence of the adaptive filter.

The new acoustic echo canceler has overcome both these problems by taking an entirely new approach to eliminating acoustical echoes. This involves adding noise-like training signals to the speech before propagation into the acoustical environment. A correlation processor at the output of the microphone will determine the reverberation modes of the environment, allowing cancellation of the echoes. A major feature of the new technique is that it is nearly independent of the presence of near-end speech.

1051-2004/96 $12.00

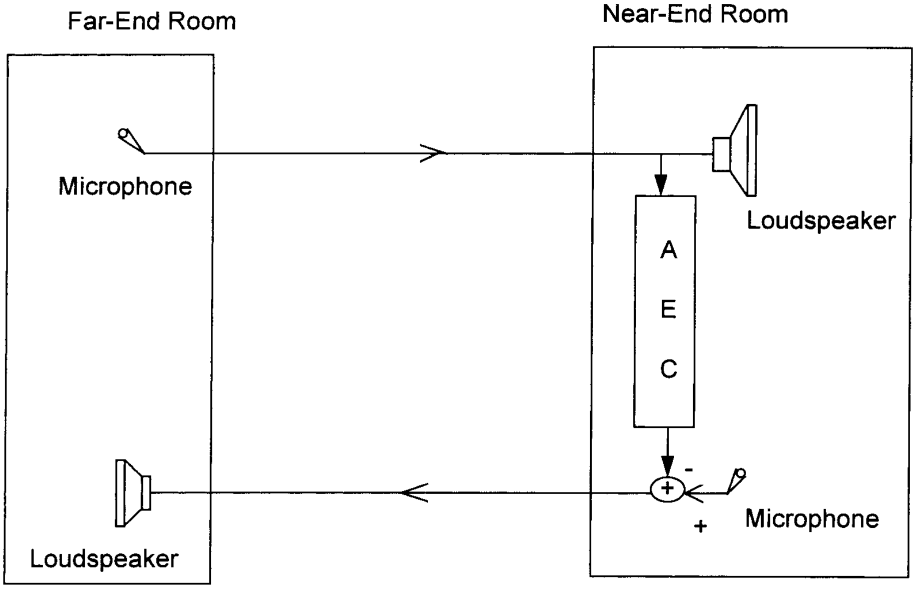

FIG. 1. A teleconference setup with an acoustic echo canceler.

Thus, the echoes are canceled and the near-end speech is transmitted undistorted. The speed of convergence of the adaptive algorithm is improved by using a filter update that is orthogonal to the input data vector. A threshold function is used to switch between the adaptive filter and the correlation estimate. These algorithms are described in the next section.

II. NEW ACOUSTIC ECHO CANCELER

The concept of an echo canceler is to consider the echo path as a mapping function which is then replicated to synthesize the echo received by the microphone. These signals are then subtracted to obtain the near-end speech alone [1]. The linear filter used to model the room impulse response can be characterized as a sequence of impulses, each completely described by two parameters: (1) a delay proportional to the delay between the speaker and the microphone and (2) an amplitude related to that delay. A linear transversal (tapped-delay-line) filter is the chosen realization for the echo-canceling filter. If the impulse response of the transversal filter is a good estimate of the room impulse response, the echo will be canceled. The output of the transversal filter is given by

$$\hat{d}(n) = \sum_{i=0}^{N-1} h_i x(n-i), \tag{1}$$

where the h_i are the coefficients of the transversal filter. These coefficients are estimated using either an adaptive algorithm or a correlation.

One of the simplest adaptive algorithms is the normalized least-mean-squares (NLMS) algorithm [2]. The implementation of such a system is not straightforward. The problems in the model result in the following requirements on the adaptive echo canceler: (1) long impulse response lengths for the linear filter, (2) fast convergence characteristics for signal inputs such as speech, and (3) fast adaptability to variations in the echo path characteristics. Another extremely important issue is the technique chosen to handle the simultaneous presence of echo and near-end speech. An unprotected adaptive filter exhibits unacceptable behavior during double talk. In [3], the performance of the NLMS algorithm for white noise and speech signals of telephone quality are discussed. It is demonstrated that the performance of the NLMS for speech input is insufficient and more sophisticated methods are required. The new acoustic echo canceler algorithm overcomes these problems.

The speed of convergence of the NLMS algorithm can be improved for speech signals if the inputs to the adaptation algorithm are decorrelated [4]. Another method for speeding up the convergence of the NLMS algorithm is based on the use of the component of the input signal vector $x(n)$ that is orthogonal to the signal vector $x(n-1)$, where n is the time index. This orthogonal component is used in the update of the adaptive filter coefficient estima-

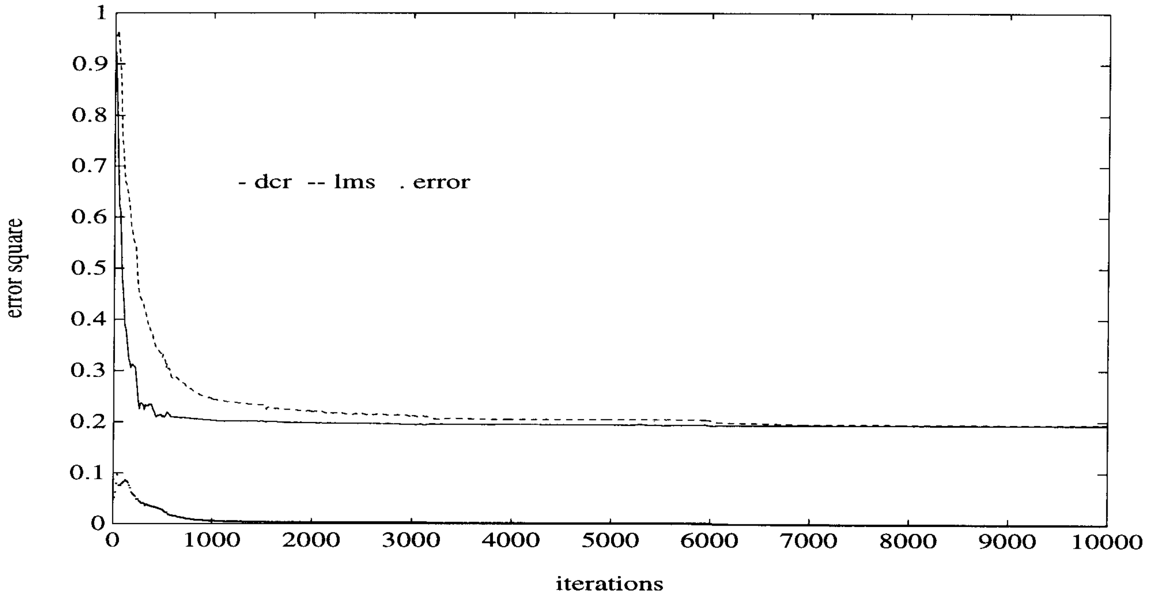

FIG. 2. Performance of the error square for the LMS and the new echo canceler algorithm.

tor. The procedure can be described as follows. The update equation is given as

$$W_N(n+1) = W_N(n) + \mu e(n) Z_N(n), \qquad (2)$$

where W are the coefficients of the adaptive estimator, μ is the step size, and $e(n)$ is the error between the signal received from the microphone and the signal estimated by the echo canceler. The equations

$$Z_N(n) = X_N(n) - c(n) X_N(n-1) \qquad (3)$$

$$c(n) = \frac{X_N^T(n) X_N(n-1)}{X_N^T(n-1) X_N(n-1)} \qquad (4)$$

are defined, where $c(n)$ is the correlation coefficient for the current input and the delayed input. This algorithm provides fast convergence of the adaptive filter for speech signals as input, and a detailed study of the algorithm along with its performance in acoustic echo cancellation is included in [5]. Performance improvements can be expected by requiring that the signal vector $Z_N(n)$ be orthogonal to the signal vector $x(n-2)$ at the expense of increased computational complexity. Taking this notion of orthogonality further approaches the recursive least-squares solution, which has a superior performance at increased computational complexity. The convergence of the new acoustic echo canceler is compared to the LMS algorithm in Fig. 2.

A correlation technique is used to combat the effect of double talk in the new acoustic echo canceler. This is done by transmitting a pseudo-noise training signal. This signal is continuously transmitted into the room and the signal received by the microphone is then correlated to estimate the filter coefficients needed to model the room. This training signal is independent of either the near-end or far-end speech and is used by the correlation algorithm to continuously and rapidly estimate the acoustic characteristics. Echo cancellation is again accomplished by subtracting the estimate of the echo signal from the microphone signal. Since the pseudo-noise training signal is uncorrelated with the other signals, it is not affected by the double talk. This is represented by the correlation process as

$$\hat{h} = \langle s_m c \rangle, \qquad (5)$$

where $\langle \quad \rangle$ represents averaging, and is defined by the equation

$$\langle x \rangle = \frac{1}{N} \sum_{i=0}^{N-1} x_i, \quad N \to \infty, \qquad (6)$$

where s_m is the signal received by the microphone and c is the pseudo-noise training signal. Expanding the signal representation and grouping terms gives

$$\hat{h} = \langle s_n c + (c * h) c + (s_f * h) c \rangle \qquad (7)$$

$$\hat{h} = h + \langle s_n c \rangle + \langle c (s_f * h) \rangle. \qquad (8)$$

When the training signal correlation with the near-end speech and the echo signal is zero (both good approximations) the estimate of the filter coefficients becomes exact:

$$\hat{h} = h. \tag{9}$$

Here s_n is the near-end speech signal, s_f is the far-end speech signal, and h is the actual impulse response of the room. We see that the correlation technique provides good estimates of the room impulse response in the presence of double talk. A detailed study of the theory of this method along with simulations of its performance in acoustic echo cancellation is included in [6].

A threshold function is used to switch between the filter coefficients obtained from the adaptive algorithm and those obtained using the correlation method. The use of the switching technique was decided upon because the adaptive algorithm performance is superior to the correlation performance when double talk is not present. Likewise, the correlation method is superior when double talk is present. The threshold function is obtained from the energy of the signal received by the microphone. A threshold level was determined from experiments conducted to detect the presence of double talk [5]. The threshold is calculated using the relation

$$t(n) = \text{abs}\left[\frac{c*e'}{e*e'}\right]. \tag{10}$$

Here c is the vector formed of the training signal send into the room and e is a vector formed of the error samples after echo cancellation. These vectors were determined experimentally to be 10 samples long. The value of $t(n)$ is bounded to an upper value of 1 and takes on values below 0.5 in the presence of double talk. This threshold level is used to switch between the adaptive and correlation algorithms.

The computational requirements of implementing this algorithms is discussed for a digital signal processor which typically does a MAC (multiply and accumulate) in one instruction. The division operation is processor dependent and is not included in the MIPS (million instructions per second) specification. The update equation (2) requires $2N + 1$ instructions, where N is the length of the echo canceler taps. Equations (3) and (4), which form the update vector, require $6N + 1$ instructions and 1 division. The correlation estimate requires NM instructions, where M is the length of the pseudo-noise training signal. The threshold detector requires 20 instructions and 1 division. It is seen that the new algorithm has inherited the low computational complexity ($O(N)$) operations of the gradient-type algorithmic schemes while providing faster convergence. Thus the new acoustic echo canceler can be implemented using 16 MIPS on a digital signal processor to obtain a 256-tap acoustic echo canceler with performance levels that approach those used by recursive least-squares algorithms which have a computational complexity of ($O(N^2)$) operations. The computational complexity of this algorithm can be reduced by efficient implementation and practical approximations. Frequency domain, block adaptive filters, or sub-band schemes of acoustic echo cancellation provide complexity reductions. However, these methods introduce delays due to block processing and their tracking ability is limited. The important note however is that these techniques cannot overcome the double-talk problem associated with acoustic echo cancellation.

III. TEST FACILITY

A working prototype of the new acoustic echo canceler was developed in the laboratory. The test bed incorporates a 486-based personal computer (PC), a pseudo-random noise generator, a set of mixers, an audio amplifier, a data acquisition board, and a digital signal processor (DSP). This test facility is shown in Fig. 3. The PC is used to store the source of the algorithms used in the acoustic echo canceler. The PC downloads these algorithms to the digital signal processor. The algorithms execute on the digital signal processor, which also controls the data acquisition board. The noise generator produces the desired pseudo-noise training signal. The tape recorder is used to provide the far-end speech. This speech is mixed with the training signal. The level of the training signal is kept well below the near-end speech level so it remains unnoticeable to the users. A second sample of the training signal is required to ensure that it is of the same level as that injected into the room. The data acquisition board samples both these signals to be used by the digital signal processor for processing. The data acquisition system mixes the speech and the training signal and provides the composite signal to an audio amplifier which drives the loudspeaker. The echo received by the microphone is sampled by the data acquisition system and provided to the digital signal processor. The amplitude of the microphone signal is amplified and controlled to provide maximum dynamic range

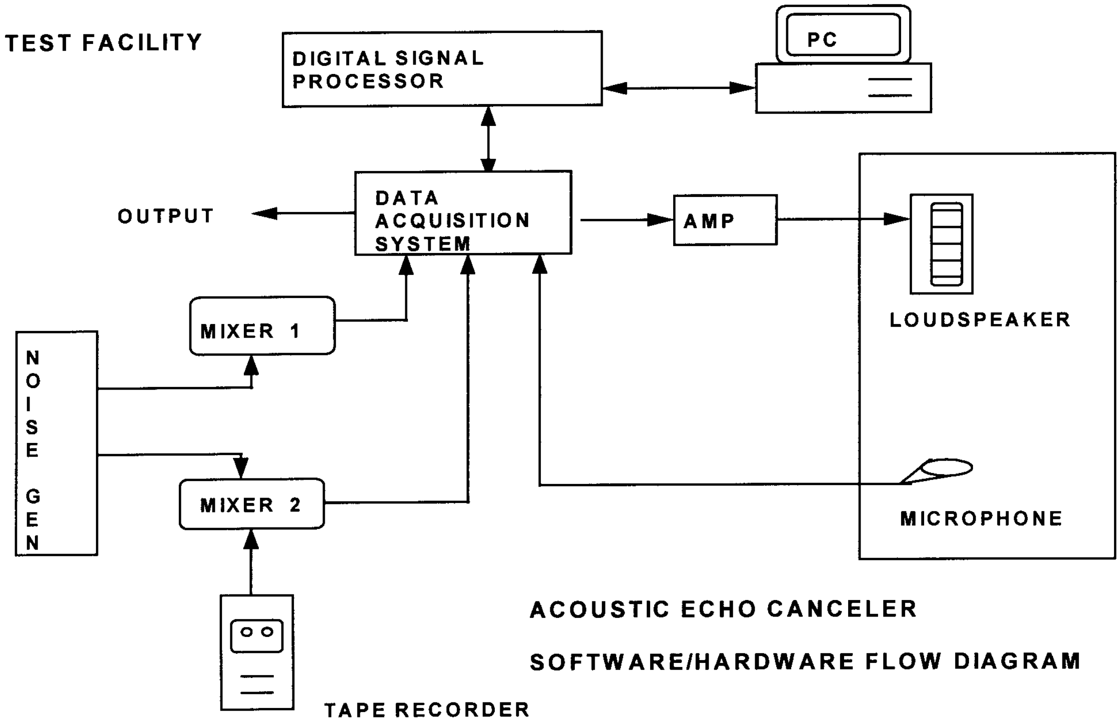

FIG. 3. Test facility for the new acoustic echo canceler.

in the analog-to-digital (A-to-D) converter. The digital signal processor uses the three signals (training signal, far-end speech plus training signal, and near-end microphone signal) to compute estimates of the filter coefficients and to do cancellation. The output of the digital signal processor is the near-end speech with the echo removed. This speech is then transmitted to the far end.

The software for the digital signal processor was written in the C language. The high-level flowchart of the main routine is shown in Fig. 4. This includes programs to: (1) implement the cancellation filter, (2) estimate the filter coefficients with an adaptive NLMS algorithm, (3) estimate the filter coefficients with the training signal, (4) calculate the threshold function value used to switch between the two coefficient estimates, and (5) control the entire system. These programs were designed, coded, and stored on the PC. For execution, the target code is downloaded to the digital signal processor. The data acquisition board consists of a four-channel analog-to-digital converter and four-channel digital-to-analog (D-to-A) converter. The sampling is done at a rate of 8000 samples per second with 16-bit resolution. The 16 bits are then transferred to the digital signal processor which carries out the processing in 32-bit floating point operations. The result is then converted to 16-bit words and transferred to the digital-to-analog converter.

In the test facility, the memory and processing speed of the hardware and research software were not sufficient to provide real-time processing. Instead, all signals were digitized, stored, and processed off-line, and the results were stored in memory. The stored results were then output to the D-to-A converter for demonstration and evaluation purposes. Extensive prototype field tests were carried out to determine the prototype performance

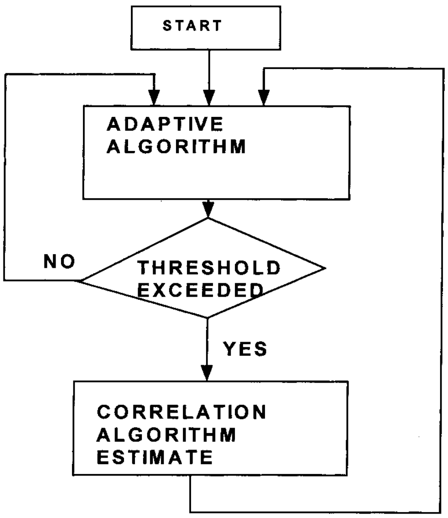

FIG. 4. Flowchart of the main routine of the acoustic echo canceler algorithm.

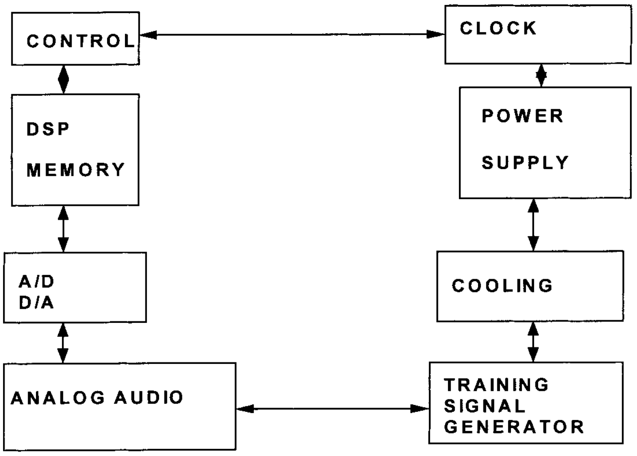

FIG. 5. Block diagram of the new acoustic echo canceler.

specifications and compare them with existing acoustic echo cancellation standards. These are summarized in the section on test results. The development of the prototype provided insight into the problems associated with the new method of acoustic echo cancellation. For example, the training signal and the speech mixed with the training signal must both be sampled at the same instant. Also, the dc offset in the analog-to-digital converter must be minimized. These were incorporated in the system design described in the next section.

IV. SYSTEM DESIGN

This section describes the design of the new acoustic echo canceler as a stand-alone unit which can be incorporated into a teleconferencing or speakerphone system. Figure 5 shows the system design.

Hardware

The analog audio includes: (1) the conditioning of analog signals, (2) the mixing of the training signal and speech, (3) the preconditioning of the signal, and (4) the audio power amplifier circuit. The analog-to-digital converter is three channels and the digital-to-analog converter is two channels. The DSP board has two fixed-point digital signal processors to carry out the computations and eight algorithm specific processors with associated memory. The control system includes the interface to the telephone lines and to the performance display. The training signal generator produces the pseudo-random noise and is driven by the clock. The clock provides synchronization to all the components. It is important to synchronize the entire system to make sure that all sampling is done at the proper time. Clock offsets anywhere in the system will lead to degraded performance.

Figure 6 shows the hardware setup of the new acoustic echo canceler. A passive backplane is used to provide support, power supply, and transmission medium for communication between the various boards. The audio board is a full PC card which contains all the audio interface and signal conditioning circuits. The 4-channel analog I/O board contains the A-to-D and D-to-A and interfaces to the DSP board. The DSP board interfaces to the analog I/O and to the adaptive filter module used to carry out the filtering. The CPU board contains a Flash EPROM and can be programmed through an RS232 interface to the PC. The entire unit can operate as a stand-alone unit and has a telephone interface to facilitate its integration into existing teleconference or speaker phone systems.

Software

The software for the system consists of modules designed for the digital signal processor. This program is used to control the data acquisition and the adaptive filter module. After the development of the desired algorithm, it is cross-compiled and downloaded to the Flash EPROM on the CPU board. This makes the unit operate as a stand-alone while providing good flexibility for software maintenance and enhancement.

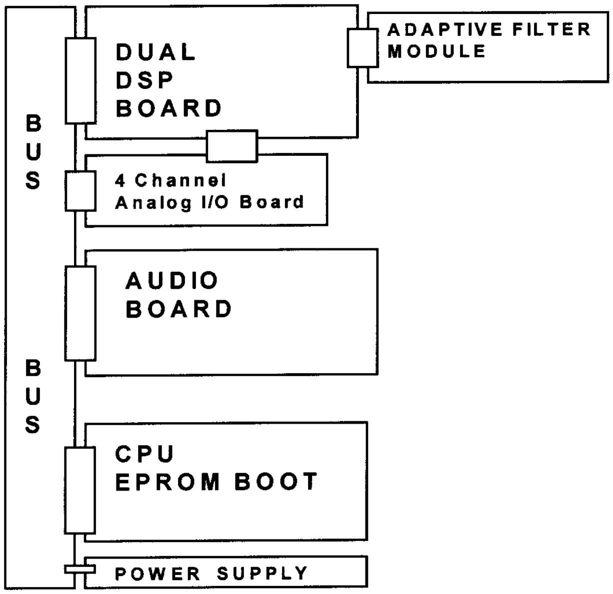

FIG. 6. Hardware setup of the new acoustic echo canceler.

V. TEST RESULTS

The two measures used to describe the performance of the echo canceler are the echo return loss (ERL) and the echo return loss enhancement (ERLE), which are measures of the amount of echo removed by the echo canceler [7,8]. These two measures are defined as

$$\mathrm{ERL}(k) = 10 \log \frac{E[x(k)x(k)]}{E[e(k)e(k)]} \tag{11}$$

$$\mathrm{ERLE}(k) = 10 \log \frac{E[y(k)y(k)]}{E[e(k)e(k)]}, \tag{12}$$

where $E[x(k)x(k)]$, $E[e(k)e(k)]$, and $E[y(k)y(k)]$ are the expected values of the far-end speech power, the expected value of the power of the uncancelled echo, and the expected value of the power of the signal received by the microphone, respectively. These terms are obtained by short-term averaging.

The implementation goal is to achieve 30 dB of ERLE because the ambient noise that is not created by the echo itself is typically measured at −30-dB level from the maximum received line signal level. The prototype achieves an ERLE value of 30 to 35 dB and an ERL value of 35 to 40 dB with no double talk. In the presence of double talk an ERLE value of 25 to 30 dB is achieved.

When first used, the acoustic echo canceler needs less than a 3-s initialization period (training) before the audio system can work properly for full duplex operation. The exact value of initialization time will depend on the characteristics of a specific room. The prototype achieves 35-dB echo cancellation in 1 s after system initialization for teleconference applications. Similarly there is an initialization time associated with the double talk. The initialization time required by the correlation algorithm is less than 10 s for teleconference applications. For the conventional acoustic echo canceler the initialization time is indefinite since the canceler cannot adapt to double talk. A recovery time is defined as the time taken by the acoustic echo canceler to achieve a steady-state ERLE value of 35 dB following a period of double talk. The prototype has a recovery time of less than 0.2 s.

The prototype acoustic echo canceler provides 35-dB echo cancellation within 1 s after system initialization and the double talk algorithm provides 25- to 30-dB echo cancellation during double-talk periods with an initialization time of less than 10 s. A recovery time of 0.2 s is needed after a double-talk period to achieve a steady-state value of 35 dB.

VI. CONCLUSION

The new acoustic echo canceler provides excellent echo cancellation under all practical acoustic operating conditions. This new method will provide means for “comfortable” hands-free telephone conversation for speaker phone and teleconference systems.

ACKNOWLEDGMENTS

This work was supported by a grant from the U.S. Department of Commerce, under USDOC Grant ITA87-02, through the Center for Advanced Technology Development (CATD) at Iowa State University, Ames. Laboratory facilities were provided by the Communication Research Group in the Department of Electrical Engineering and Computer Engineering.

REFERENCES

1. Gritton, C. W. K., and Lin, D. W. Echo cancellation algorithms. *IEEE ASSP* (Apr. 1984).
2. Haykin, S. *Adaptive Filter Theory.* Prentice–Hall, Englewood Cliffs, NJ, 1991, 2nd ed.
3. Hansler, E. The hands-free telephone problem—An annotated bibliography. *Signal Process.* **27** (June 1992), 259–271.
4. Ozeki, K., and Umeda, T. An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties. *Electron. Comm. Jpn.* **67-A** (1984), 19–22.
5. Porayath, R. Adaptive algorithm for acoustic echo cancellation, M.S. thesis, Iowa State University, Ames. Dec. 1993.
6. Miller, J. W. “Correlation-based acoustic echo canceler,” M.S. thesis, Iowa State University, Ames. Dec. 1993.
7. Park, S., and Messer, D. Acoustic echo canceler for full-duplex voice transmission on fading channels. In *Proc. International Mobile Satellite Conference, Ottawa, Canada,* June 17–20, 1990, pp. 188–193.
8. Park, S., and Hillman, G. On acoustic echo cancellation implementation with multiple cascadable adaptive FIR filter chip. In *Proc. Int. Conf. Acoust., Speech, Signal Processing, Glasgow, Scotland,* May 1989, pp. 952–955.

RAJIV PORAYATH was born in Palghat, India. He received his B.E. degree in 1990 from Manipal Institute of Technology, Manipal, India. His M.S. was received in 1993 from Iowa State University, Ames. From 1992 to 1994, he was a research assistant, working on acoustic echo cancellation, in the Department of Electrical and Computer Engineering, Iowa State University. He worked on acoustic echo cancellation problems as an assistant scientist with the Center for Advanced Technology Development (CATD) in Ames, Iowa. He is currently working with Shure Brothers, Incorporated, Evanston, IL. He is a member of Eta Kappa Nu and the IEEE.

JOHN F. DOHERTY received the B.S. (Hon) degree from the

City University of New York in 1982, the M.Eng degree from Stevens Institute of Technology, Hoboken, NJ, in 1985, and the Ph.D. degree from Rutgers University, New Brunswick, NJ, in 1990, all in Electrical Engineering. He worked as an integrated circuit reliability engineer at IBM from 1982 to 1984. He was a member of the technical staff at AT&T Bell Laboratories from 1985 to 1988, working in the area of underwater acoustical signal processing. Currently, he is an assistant professor and a Harpole–Pentair Fellow at Iowa State University, Ames, in the Electrical Engineering and Computer Engineering Department. His current research activities include interference rejection in CDMA systems, hot clutter suppression in airborne pulsed radar, and ECG signal analysis for tachycardia risk assessment.

STEVE F. RUSSELL was born in Lewiston, Idaho. He received a B.S. degree in 1966 from Montana State University, Bozeman. His M.S. and Ph.D. in Electrical Engineering were received in 1973 and 1978, respectively, from Iowa State University, Ames. From 1966 to 1970, he was in the Telecommunications Division at Collins Radio Company. In 1976, he joined the Avionics Division of Rockwell–Collins. From 1980 to 1984, he was a principal engineer at King Radio Corporation. He joined the faculty of the Department of Electrical Engineering and Computer Engineering at Iowa State University as an associate professor in 1984. His research has been in the areas of spread-spectrum ultrasonic testing, spread-spectrum acoustic echo cancellation, complex FIR digital filter design, frequency-hop communications intercept, HF networking, and radiograph image processing.